

# La Iniciativa de Archivos Abiertos (OAI), un nuevo paradigma en la comunicación científica y el intercambio de información

Laureano Felipe Gómez Dueñas\*

## RESUMEN

Las bibliotecas, archivos, centros de documentación y todas las áreas que trabajan con información, están pasando por una etapa de evolución que involucra un manejo oportuno y eficiente de la información digital. La iniciativa de archivos abiertos ha surgido como una alternativa para solucionar muchos de los problemas que hoy enfrentamos cuando trabajamos con sistemas de información y con la información digital contenida en estos; adicionalmente sirve como herramienta de apoyo a los movimientos organizados por los científicos y bibliotecólogos de todo el mundo que buscan liberar los contenidos científico-académicos y dejarlos al alcance de la humanidad, también brinda una alternativa de trabajo ante la dificultad de integración entre los distintos sistemas de información documental y la poca normalización y aplicación de estándares entre estos.

**Palabras clave:** iniciativa de archivos abiertos, bibliotecas digitales, metadatos, *Dublin Core*, interoperabilidad, documento digital, colecciones digitales, repositorios institucionales, auto-almacenamiento.

## THE INITIATIVE OF OPEN ARCHIVES (OAI), A NEW PARADIGM IN THE SCIENTIFIC COMMUNICATION AND THE INFORMATION INTERCHANGE

### ABSTRACT

Libraries, archives, documental centers, and all the areas that deal with information are going through a stage of evolution that involves the opportune and effective use of digital information. The open archives initiative arises as an alternative to solve most of the problems currently faced when working with systems of information and their digital data. Furthermore, these achieves work as a supporting tool for the scientist and librarian worldwide organizations that promote the free access to scientific and academic contents for everyone. It also provides an alternative against the low integration amongst the different systems of documental information and the little standardization as well as its implementation within them.

**Keywords:** initiative in open files, digital libraries, Dublin Core, interoperability, digital document, digital collections, institutional repository, self-filing.

\* Ingeniero de Sistemas de la Universidad Nacional de Colombia. Especialista en Redes Documentales de la Pontificia Universidad Javeriana. Correo electrónico: felipe.gomez3@unisabana.edu.co

## INICIATIVA DE ARCHIVOS ABIERTOS

Vivimos actualmente en la denominada «sociedad de la información», caracterizada principalmente por una gran explosión de la información digital que se ha originado por la rápida evolución de las tecnologías, el auge de Internet y la oportunidad de acceso y socialización de éstas. Sin embargo, esta rápida evolución ha traído nuevos problemas y oportunidades a los profesionales de la información, ya que se ha generado un gran desorden documental ante la inmensidad de información existente, la gran duplicación de datos y la poca comunicación e interacción de los sistemas de información que la contienen.

La Iniciativa de Archivos Abiertos, se desarrolló como una alternativa que busca promover el libre acceso a la información científica de manera ordenada y normalizada (Lynch, 2001) mediante la implantación en los repositorios institucionales de documentos y bibliotecas digitales existentes en el mundo, de unas metodologías comunes que persiguen estandarizar la interoperabilidad entre éstos, con el fin de facilitar de forma eficiente y transparente la diseminación de contenidos mediante el desarrollo e implantación de una metodología y protocolos sencillos y prácticos que permiten una rápida aplicación de las normas y estándares internacionales relacionados con manejo de información digital.

La Iniciativa de Archivos Abiertos (OAI), tal como se ha venido desarrollado en el mundo, será la responsable que permitirá crear un gran sistema distribuido de documentos digitales (Biblioteca Digital Mundial), el cual se ha estado formando gradualmente mediante el proceso de compartir los metadatos existentes en distintos repositorios de documentos y bibliotecas digitales, distribuidas

éstas geográficamente por todo el mundo que probable, pero no necesariamente, manejan sus documentos de forma estandarizada (en forma y estructura), cuya información puede no ser indexada por los buscadores tradicionales y por lo tanto ser opaca para la mayoría de las comunidades académicas y científicas. Esta integración se logrará a través de la implementación del protocolo de recolección (cosecha) de metadatos *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*, la cual se consigue mediante unos procedimientos que generalmente son rápidos de realizar y de muy bajo costo económico (Van de Sompel y Lagoze, 2000).

## PROBLEMAS EN EL MANEJO DE LA INFORMACIÓN DIGITAL

El manejo y control de la información digital existente en la actualidad, ha generado nuevos retos y problemas que requieren ser tratados y solucionados de forma inmediata para así poder utilizar ésta de forma óptima. La Iniciativa de Archivos Abiertos (OAI) busca crear soluciones y oportunidades mediante procedimientos prácticos y sencillos a los siguientes problemas:

### PRIMER PROBLEMA: CALIDAD Y PERTINENCIA EN LA RECUPERACIÓN DE INFORMACIÓN

El primer problema está en la gran cantidad de información existente en Internet, la dificultad de acceder a información relevante y de calidad en un contexto donde al realizar una búsqueda en un portal especializado como Google, nos arroja miles de enlaces. Google ha reportado que en el año 2002 recibió más de 55 billones de búsquedas y actualmente indexa más de ocho billones (8.058.044.651) de páginas Web y documentos en texto completo (PDF, DOC, PPT, etc..) (Google, 2003), aunque este motor de búsqueda tiene un sistema de evaluación

de relevancia muy especializado (utilizando los algoritmos de búsqueda como *Page Ranking*, Indización Semántica Latente, entre otros) (Martínez, 2005), en esencia buscan que el sistema intente proveer los mejores resultados infiriendo una relación con la pregunta en su contexto propio, es decir, que se aleja de la premisa del «término exacto» estrechamente vinculada a los operadores *Booleanos* tradicionales, aún así, si no utilizamos los operadores y propiedades avanzadas de búsqueda, el 80% de la información encontrada no coincidirá con nuestra necesidad de información (generalmente los primeros enlaces son de información publicitaria o empresas comerciales que aprovechan las formas de indización de estos buscadores para aparecer en las primeras posiciones) o simplemente los documentos encontrados contendrán información desarrollada sin ningún carácter científico o metodológico; esto nos introduce una gran demora en obtener información que nos pueda aportar en la elaboración de nuevo conocimiento. Parte de este problema se debe a la desorganización con la que fue creado Internet y la ausencia de un mecanismo de integración de metadatos normalizados junto a la información presentada en los documentos *Web*. En este momento aún desconocemos el gran poder de la *Web* invisible, miles de bases de datos de bibliotecas y centros de información especializados, muchos de ellos con información en texto completo que por no poseer una efectiva forma de diseminación y/o de compartir sus recursos, se pierden inevitablemente. Google estima que pese a su gran algoritmo de búsqueda e indexación de información, apenas alcanza a explorar el uno por ciento (1%) del total de los recursos de información existentes en línea (Bergman, 2001). En la búsqueda de una solución, la iniciativa de archivos abiertos busca crear un gran sistema de conocimiento científico universal, con los metadatos de toda la información académica existente (referencia a los documentos físicos y preferencialmente

con disponibilidad de texto completo en línea) utilizando todos los tipos de programas y plataformas existentes sin distinción, colocando así información de relevancia en Internet, y consolidando un sistema de búsqueda que compita con los grandes motores de búsqueda del mundo (OAI vs Google).

## SEGUNDO PROBLEMA: EDITORIALES VS *OPEN ACCESS*

El segundo problema se concentra en las facilidades para la difusión y la accesibilidad de la información científica a través de Internet. Éstas contrastan con las barreras tanto económicas como de derechos de producción (*copyright*) impuestas por las grandes editoriales que controlan la mayoría del mercado de las publicaciones científicas (Melero, 2004). La iniciativa de acceso abierto de Budapest (*Budapest Open Access Initiative*), es el primer eslabón dentro de un movimiento internacional organizado por los científicos, académicos y bibliotecólogos de todo el mundo que buscan liberar todos los contenidos y dejarlos libres al alcance de la humanidad. Esto ha causado un cambio en la mentalidad mundial acerca de la publicación de contenidos, ya que los creadores de información y conocimientos desean obtener de forma más fácil y en el menor tiempo posible, mayor prestigio y reconocimiento por sus investigaciones. «Los artículos que están libremente expuestos en la red, son mucho más citados que aquellos que están escondidos en los arcanos del papel impreso; estos artículos *online* son citados 4.5 veces más frecuentemente que los que no están en línea» (Lawrence, 2001).

Esta conclusión nos indica que si se desea tener mayor impacto científico, si se quiere ayudar a la circulación de ideas y si se busca compartir y potenciar el conocimiento generado en las universidades, laboratorios y centros de investigación, es obligatorio que la información

circule lo más libremente posible en contradicción con los intereses de las grandes editoriales, de las universidades privadas, y sobretodo de las empresas editoriales que sobreviven de la gestión del conocimiento (Lawrence, 2001).

Adicionalmente se está evidenciando actualmente que las universidades iberoamericanas, deseosas de que se conozcan rápidamente sus investigaciones y adelantos realizados, están haciendo esfuerzos individuales, grupales y por país, muchas veces gracias a la ayuda de organizaciones internacionales (UNESCO, OPS Bireme - Scielo, Portal Tecnociencia), para colocar los contenidos en texto completo de sus publicaciones en línea y de libre acceso, logrando así un reconocimiento y mayor difusión internacional, más allá del simple interés económico de captar suscripciones.

**Proyecto Scielo:** Scielo - *Scientific Electronic Library Online* (Biblioteca Científica Electrónica en Línea) es un modelo para la publicación electrónica cooperativa de revistas científicas en Internet. Especialmente desarrollado para responder a las necesidades de la comunicación científica en los países en desarrollo y particularmente de América Latina y el Caribe. El modelo proporciona una solución eficiente para asegurar la visibilidad y el acceso universal a su literatura científica, contribuyendo a la superación del fenómeno conocido como «ciencia perdida». Además, el modelo Scielo contiene procedimientos integrados para la medida del uso y del impacto de las revistas científicas, también contempla el desarrollo de una metodología común para la preparación, almacenamiento, diseminación y evaluación de la literatura científica en formato electrónico. El proyecto Scielo nació como una iniciativa de FAPESP (Fundación de Apoyo a la Investigación del Estado de São Paulo<sup>1</sup>) y de BIREME<sup>2</sup> (Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud).

**Portal Tecnociencia:** el portal tecnociencia (<http://www.tecnociencia.es/>), dirigido por la FECYT (<http://www.fecyt.es>) con el apoyo técnico y operacional del Centro de Información y Documentación Científica (CINDOC) (<http://www.cindoc.csic.es>), ha puesto en marcha un portal de revistas electrónicas de libre acceso y a texto completo, llamado «e-revistas» (<http://www.tecnociencia.es/e-revistas/>). Esta plataforma digital cuenta con la participación, hasta el momento, de 45 revistas españolas de todas las áreas científicas. Sus objetivos son:

- ◆ Establecer un sistema de calidad normalizado y validado de las revistas electrónicas.
- ◆ Crear una plataforma que aglutine a todas las revistas electrónicas que quieran formar parte de la misma y cumplan los requisitos de calidad establecidos.
- ◆ Crear una interfase OAI-PMH (*Open Archives Initiative - Protocol for Metadata Harvesting*) que permita que los contenidos del portal de revistas estén disponibles en múltiples bases de datos y catálogos internacionales.
- ◆ Aportar servicios de valor añadido sobre los documentos indicados.
- ◆ Difundir el movimiento de acceso abierto entre la comunidad científica nacional.
- ◆ Asesorar técnicamente para la creación de nuevas revistas electrónicas de acceso abierto.
- ◆ Proporcionar servicio opcional de hospedaje de revistas electrónicas de libre acceso.
- ◆ Difundir ampliamente el proyecto entre instituciones académicas para su conocimiento y uso de los recursos generados.

Como solución a este problema, la Iniciativa de Archivos Abiertos (OAI) busca ser una herramienta que contribuya significativamente a abreviar el ciclo de vida de la comunicación científica, brindando

1 <http://www.fapesp.br>

2 <http://www.bireme.br>

acceso inmediato, libre y global a la información mientras ésta se va produciendo en la medida que se posibilite, por ello se ha dividido esta información en tres grandes grupos (Hunter y Guy, 2004):

- ◆ **Pre-Prints:** este significa un documento que no ha sido aún avalado por pares académicos (*non-peer reviewed document*), que está disponible en medio digital antes de su evaluación y publicación impresa.
- ◆ **E-Prints:** corresponde a los documentos en estado terminado y lista para su publicación (generalmente avalada por pares académicos), cuyos autores desean que esté disponible inmediatamente en Internet y de acceso libre.
- ◆ **Post-Prints:** son los documentos que ya se han publicado en alguna revista con cierto tiempo de edición, y cuya información se puede colocar libremente en Internet.

El ideal es que esta información se encuentre de libre acceso, sin más restricciones al *copyright* que la de ser debidamente citados, generalmente en repositorios especializados de *pre-prints*, *e-prints* ó *post-prints*. Asimismo las grandes casas editoriales (Elsevier-ScienceDirect, Ingenta, Springer, etc.), las editoriales universitarias y las organizaciones que apoyan el acceso abierto (BioMed Central, PubMed, DOAJ, Free Medical Journals), están haciendo esfuerzos particulares por implementar el protocolo de recolección de metadatos de la iniciativa de archivos abiertos (OAI-PMH) en sus sistemas de información propios, y lograr con esto una mayor difusión de sus contenidos (algunos gratuitos), aprovechando que los grandes motores de búsqueda de Internet (Scirus, Google, Yahoo, etc..), ya están indexando este tipo de información por medio de la implementación de los preceptos de esta iniciativa.

**¿Cómo estamos en Colombia?** En Colombia, en el ámbito académico se están manejando actualmente

dos iniciativas: La consolidación de portales de revistas (Encolombia.com, Scielo Colombia, etc...) y la publicación independiente en las universidades de sus contenidos utilizando diversas plataformas de publicación tanto propietarias como de software libre.

**Portal [www.encolombia.com](http://www.encolombia.com) (<http://www.encolombia.com/encolombia.htm>).** En este portal, se encuentran 21 revistas en medicina y 11 revistas de odontología, cuyo contenido se puede acceder en texto completo, apoyadas por anuncios muy grandes de productos farmacéuticos.

**Portal Scielo Colombia (<http://scielo-co.bvs.br/scielo.php>).** En este portal, se encuentran en texto completo 65 números de 7 revistas especializadas en medicina, actualizados permanentemente por sus editores.

Con el desarrollo de la Iniciativa de Archivos Abiertos (OAI), se brindan los mecanismos técnicos y se pone de manifiesto que la disponibilidad libre de documentos en línea facilita el acceso en múltiples modos, desde la obtención de los archivos, pasando por la conexión directa entre los científicos (Correo electrónico, conversaciones, foros y videoconferencias por Internet), además de la creación de grupos de trabajo y de discusión, la indexación por medio de motores de búsqueda en Internet y la creación de nuevos servicios tecnológicos.

### TERCER PROBLEMA:

#### INTERACCIÓN ENTRE SISTEMAS DE INFORMACIÓN

El tercer problema está en la disparidad de la información y los sistemas informáticos de gestión de ésta en las organizaciones (bibliotecas, centros de documentación, archivos y otros), cuyos sistemas de información fueron diseñados generalmente para atender necesidades particulares, sin seguir nin-

guna norma o estándar, y que por esto no se pueden interconectar ni compartir su información plenamente. Existe un problema profundo a nivel interno de heterogeneización de productos de *hardware* y *software*, además de la forma de ingreso de la información (niveles de catalogación), ya que existen muchos datos repetidos, dispares y contradictorios si comparamos cada sistema existente.

Pese a lo anterior, existe el interés de las bibliotecas en compartir su información mediante la integración de sus sistemas y catálogos bibliográficos, con el ánimo de dar a conocer sus colecciones y servicios de información de una manera fácil y eficiente.

Según Cole (2003) los bibliotecólogos asumen que una copia de casi cada libro impreso reside en por lo menos una biblioteca o archivo en alguna parte en el mundo, y ellos sienten que como especialistas de información lo deben poder encontrar fácil y rápidamente. Sin embargo, las principales soluciones hasta ahora propuestas de generar catálogos colectivos y unir las colecciones de información existentes han encontrado grandes tropiezos en cumplir su misión:

- ◆ La primera solución probada, consiste en que cada institución enviara a un nodo principal su información existente, sin importar el estado de ésta, su formato y tipo. Lo importante era recoger la información, el nodo principal se encargaría de «traducir» la información obtenida a un modelo normalizado, y por medio de un programa especializado, incorporarla a un sistema centralizado de información en el cual todos los usuarios pudieran consultar todas las colecciones existentes en los nodos cooperantes. Éste modelo tiene como ventaja la gran velocidad de respuesta al momento de realizar una consulta, ya que la información se encuentra

local en el servidor, y en el aumento en la pertinencia de la información buscada, ya que solamente se requiere aprender un modo de interrogación del motor de búsqueda para obtener la información de todos los nodos. Entre las desventajas podemos citar principalmente:

- ◆ Cualquier cambio al formato y/o tipo del documento, requeriría rediseñar y cambiar el programa traductor de información, lo cual conlleva altos costos de programación.
- ◆ Evitar la duplicidad de la información, aun cuando ésta venga de la misma fuente, requiere altos costos de procesamiento, ya que no hay control adecuado de la información suministrada por cada nodo. En el momento de importar la información al sistema central, éste tiene que verificar que dicha información no se haya registrado anteriormente por el mismo u otro nodo y debe poder discernir la información nueva de la presente.
- ◆ Añadir un nuevo nodo requiere un nuevo trabajo de traducción de la información enviada.
- ◆ La segunda propuesta consiste en la utilización del protocolo Z39.50, el cual es un protocolo de comunicación estándar que se usa para la comunicación transparente entre sistemas de información bibliotecarios; éste trabaja en una arquitectura cliente servidor basada en estados y es una norma (ANSI Z39.50 / ISO 23950). Z39.50 es un protocolo dirigido a facilitar la búsqueda y recuperación de información en distintos sistemas a través de una misma interfaz. Su aplicación en el mundo de las bibliotecas y de los centros de documentación permite la consulta de recursos distribuidos en distintas bases de datos desde un mismo punto

de acceso. La creación de catálogos colectivos virtuales o la cooperación en las tareas de catalogación son algunas de las ventajas que ofrece este protocolo a los profesionales de la información (Sedic, 2005). Sin embargo, la difusión en Colombia de este protocolo se ha desarrollado muy lentamente debido al problema básico de la dificultad de su interpretación, desarrollo e implementación en cualquier sistema de información por parte de los ingenieros programadores. El protocolo Z39.50 tiene una funcionalidad muy completa, ya que trabaja con manejadores de sesión y activadores de resultados para crear servicios de alerta y disseminación selectiva de información; además, permite una especificación de predicados que filtran los registros retornados de una búsqueda. El utilizar estas funcionalidades viene a incrementar la dificultad de implementación y los costos asociados. Debido a la especificidad del protocolo y la gran cantidad de normas que éste involucra, no todos los sistemas bibliográficos contemplan la implementación de un servidor Z39.50. Adicionalmente, otro problema que nace de la utilización de este protocolo es el consumo de una gran cantidad de recursos de red ocasionado por cada búsqueda, y el tiempo de respuesta de estas búsquedas distribuidas es alto cuando se realiza en más de 20 nodos simultáneamente (Piñeiro y Marmonti, 2004)<sup>3</sup> ó cuando algún nodo de la red no contesta rápidamente por estar desconectado o apagado. De lo anterior podemos deducir que aún no existen tecnologías adecuadas para lograr la interoperabilidad a bajo precio utilizando el protocolo Z39.50.

Una solución que nos brinda la Iniciativa de Archivos Abiertos (OAI), es que podemos combinar los dos modelos anteriormente presentados (base de datos centralizada, sistemas de información distribuidos). OAI nos propone una solución de integración e interoperabilidad de repositorios de información, para la creación de catálogos colectivos, mediante la utilización práctica de normas y estándares internacionales con una metodología de fácil uso y bajo costo de implementación, esto a través del proceso de recolección (cosecha de metadatos), en la cual periódicamente, los sistemas de información distribuidos se comunican y se actualizan entre sí de forma automática, y se genera un sistema centralizado para búsqueda y recuperación de información.

### LA INICIATIVA DE ARCHIVOS ABIERTOS (OPEN ARCHIVES INITIATIVE - OAI)

*Open Archives Initiative* (OAI) es una iniciativa apoyada financieramente por la Federación de Bibliotecas Digitales (DLB) y la Liga para la Información en Red (CNI) y la Fundación Nacional para el Desarrollo de la Ciencia (NSF), que desarrollan y promueven la aplicación de estándares de interoperabilidad para facilitar la eficiente disseminación de contenido (Gómez y Arias, 2002).

Esta iniciativa, nació como una respuesta a la necesidad de los investigadores de poder crear una red académica en la cual se pudiese conocer acerca de todas las investigaciones existentes en el mundo que, además de permitir ubicar información académica de calidad en un solo punto, se tuviera disposición a obtener el material en texto completo de forma inmediata y libre de cargos. Esto permite una mayor comunicación entre los proyectos existentes y la duplicación de esfuerzos.

3 PIÑEIRO, ISABEL y MARMONTI, EMILIANO. Metadatos una introducción. *II jornadas sobre la Biblioteca Digital Universitaria* (<http://www.econ.uba.ar/www/servicios/Biblioteca/ponencias/metadatos%20marmonti%202.ppt>).

Los principios básicos de esta iniciativa estriban en los conceptos de interoperabilidad, auto-almacenamiento y el libre acceso para la revisión por la comunidad (Triska y Café, 2001).

Podemos definir interoperabilidad como la capacidad de un sistema de información de comunicarse y compartir información efectivamente con otro mediante una interconexión libre y transparentemente (compartir metadatos, documentos y objetos digitales), en términos del OAI se busca crear una metodología minimalista y simple que esté al alcance del personal técnico y bibliotecológico, con una forma de entendimiento rápida, que sea de fácil desarrollo y aplicación (contrario a Z39.50), que a partir de la masificación del uso del protocolo HTTP, el manejo de páginas *Web* y el uso extendido de los navegadores de Internet, se pueda añadir a los software de sistemas de información existentes capacidades de interoperabilidad y planear en los nuevos desarrollos la utilización de estas capacidades como medio de transporte de la información, incluyendo además otros conceptos y métodos derivados como son los *Webservices* y el manejo del metalenguaje XML, los cuales se consideran actualmente como los modelos ideales para el manejo de información digital, respecto al nivel de catalogación y modelo de metadatos a utilizar, OAI se basa inicialmente en el estándar *Dublin Core* (ISO 15836: 2003), en formato simplificado (sin calificadores de campo), al ser este el modelo más fácil de implementar, el cual está diseñado para trabajar principalmente con información digital y poseer un modelo de implementación fundamentado en XML. Aunque el diseño del protocolo permite especificar cualquier modelo de metadatos existente (marcxml, rfc1807), además permite crear nuevos modelos de metadatos

mientras estos estén formulados en XML (etd-ms, mtd-br, mtd-col).

El auto-almacenamiento es una forma de agilizar el flujo de la comunicación científica, en la cual los autores utilizan su derecho de publicar su producción intelectual (total o parcialmente terminada) sin la presencia de un intermediario (editoriales), generalmente estas publicaciones se colocan como *pre-prints*, en grandes repositorios de información especializada existentes en Internet (Arxiv, E-lis, Repec, etc.), bajo esta premisa de publicación el autor espera favorecer el acceso a la mayor cantidad de personas posibles, obtener mayor reconocimiento y así mismo obtener rápidamente una retroalimentación de su trabajo. Estos archivos ó repositorios de documentos especializados proveen el libre acceso para a la comunidad enmarcado íntimamente el movimiento de acceso abierto de Budapest y otras iniciativas que han surgido como el movimiento de *copyleft* (Wu, 2003) y la licencia de publicación *Creative Commons* (Ariel, 2003)<sup>4</sup> permiten realizar búsquedas dentro de sus colecciones y además ofrecen otros servicios adicionales como la generación de alertas de nuevos documentos ingresados, estos repositorios generalmente son indexados por los grandes motores de búsqueda de Internet y así obtienen los documentos almacenados en éstos una mayor visibilidad. Cuando un autor publica su investigación en repositorios abiertos, se está solucionando en gran medida los problemas del manejo de derechos de autor, ya que el mismo creador está colocando su información en forma consciente y libre, además de tener espacios propios para el control y almacenamiento de sus documentos y versiones de éstos.

---

4 ARIEL HERNÁN VERCELLI. Creative Commons y la profundidad del copyright. *Revista Enredando*. N° 353-105° de la 4ª versión ([http://www.sindominio.net/afe/dos\\_copyleft/cc.pdf](http://www.sindominio.net/afe/dos_copyleft/cc.pdf)). Última consulta 2003.



La ventaja del nacimiento de esta iniciativa en el seno de las organizaciones científicas radica en el hecho de que siempre han trabajado con grandes cantidades de información especializada y por lo tanto han creado modelos de repositorios de documentos a medida que se requería unir y difundir la información producida por una comunidad científica o académica, y es así que se han desarrollado globalmente y gracias al gran auge de éstas actualmente se trabajan y desarrollan nuevos proyectos a nivel mundial que trabajan estrechamente la OAI, las cuales involucran proyectos de implantación de repositorios institucionales y bibliotecas digitales interoperables, además de la generación de nuevos desarrollos de *software* (generalmente *software* libre) que, basados en la utilización novedosa de este protocolo crean nuevos modelos para el manejo de información digital.

La historia de esta iniciativa (Van de Sompel y Lagoze, 2000) comienza en julio de 1999 con un Llamado de Participación emitido por Paul Ginsparg (ArXiv, s.f.)<sup>5</sup> y Rick Luce del Laboratorio Nacional de Los Álamos (s.f.)<sup>6</sup> y Herbert Van de Sompel de la Universidad de Cornell, a una reunión para explorar la cooperación entre repositorios académicos electrónicos existentes hasta la fecha, que culminó con la Convención de Santa Fé (Ginsparg *et al.* 1999), celebrada en Nuevo México en octubre del mismo año (Wu, 2003). El tema central de la primera reunión era el establecimiento de recomendaciones y de mecanismos para facilitar servicios de valor añadido mediante la integración de los repositorios existentes. Tales servicios se podrían crear a partir de la información combinada obtenida de los centros cooperantes mediante procesos automáticos, para poner esa información realzada a disposición los

usuarios, los agentes, u otros servicios. Los ejemplos de tales servicios incluyen los motores de búsqueda múltiples (búsqueda cruzada - metabúsqueda), servicios de diseminación selectiva de información, sistemas de enlace múltiple y servicios de arbitraje de documentos.

Durante esta reunión se examinaron, las experiencias, problemas y soluciones empleadas por estos repositorios, adicionalmente se analizó las distintas propuestas que existían hasta entonces para compartir la información de cada repositorio utilizando un modelo balanceado de costo/ funcionalidad, a partir de esto se determinó un modelo de interoperabilidad multicapa que buscaba crear una metodología sencilla, que fuese de fácil implantación, y creado a partir de todas las soluciones posibles existentes a la fecha (protocolos de Guilford, Dienst, Harvest, Z39.50, etc.) como un modelo mixto.

Adicionalmente a partir de las conclusiones de esta reunión se creó la expectativa de la creación de un servicio universal para el almacenamiento de literatura académica y científica (The Universal Preprint Service - UPS) (Van de Sompel *et al.*, 2000), el cual cobija la difusión únicamente de la información en texto completo y de acceso libre, este modelo es la base conceptual para el desarrollo de repositorios que manejan la iniciativa de archivos abiertos. Bajo esta premisa, se han desarrollado multitud de programas que integran el desarrollo de *software* libre, junto al movimiento de acceso abierto (*Open Access*). Inicialmente se pensó unir exclusivamente los repositorios de *e-prints*; sin embargo, el desarrollo de esta iniciativa, permite integrar cualquier sistema de información, por ello se adoptó un objetivo mucho más amplio:

5 ARXIV. Archivo y servicio de impresos electrónicos en los campos de la Física, Matemáticas y Computación. [En línea]: (<http://arxiv.org>). Última consulta 2005.

6 ALAMOS LABORATORY. El Laboratorio Nacional de Los Álamos principal centro de investigación de los Estados Unidos. [En línea]: (<http://www.lanl.gov/>). Última consulta 2005.

abrir el acceso a un rango de materiales digitales (Barruelo y Subirats, 2003) que involucraban compartir los metadatos de cualquier tipo de material que estuviese en formato digital (audio, música, videos, animaciones, planos, mundos y recorridos virtuales, objetos de aprendizaje, etc.).

## PROVEEDORES DE DATOS VS PROVEEDORES DE SERVICIOS

Todos los programas desarrollados según el modelo UPS, están divididos en dos clases de participantes distintos de acuerdo al tratamiento de la información: proveedor de datos y proveedor de servicios (Van de Sompel *et al*, 2000), la interoperabilidad entre estos dos tipos de programas es la esencia del trabajo de la OAI, el modelo no contempla ninguna particularidad respecto a base de datos, lenguaje de programación, creador, diseño, y manejo interno de la información.

**Proveedores de datos:** es básicamente un administrador de contenidos que soporta el protocolo OAI, éste fue diseñado y enfocado inicialmente a la custodia de *e-prints* que permitiera un medio de exposición de sus metadatos y de los documentos asociados. Un proveedor de datos puede construirse desde lo más básico, como crear un archivo en XML siguiendo una estructura específica, y que contiene las referencias (metadatos) de unos documen-

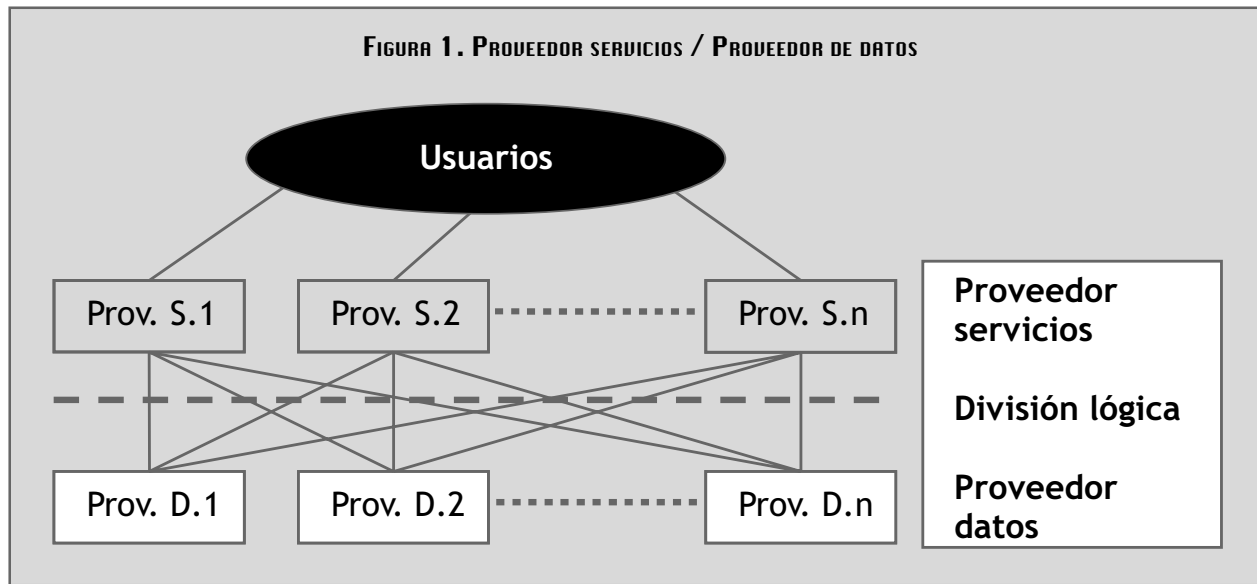
tos resguardados (acceso público o privado), pasando por los repositorios de documentos que permiten esencialmente el almacenamiento, la navegación y la búsqueda, hasta llegar a un entorno más avanzado mediante la utilización de una biblioteca digital, que incorpora además del contenido ofrecido por un repositorio común, servicios digitales de valor agregado como, por ejemplo, agentes digitales y diseminación selectiva de la información. Adicionalmente, al enfoque inicial, son tomados como proveedores de datos: una base de datos que contiene la catalogación de los materiales físicos de una biblioteca, una revista electrónica o un sistema de almacenamiento de objetos de aprendizaje, entre otros.

**Proveedores de servicios:** se puede definir como un programa encargado de recolectar la información ofrecida por los proveedores de datos (metadatos) distribuidos estos por Internet federadamente, en un sistema centralizado utilizando el protocolo OAI, a partir del cual se toma como base para construir servicios de valores agregados a los usuarios finales.

Entre estos servicios, se pueden incluir los metabuscadores, las bibliotecas digitales federadas y los localizadores de recursos en Internet (*document resolvers*). Para poder recolectar dicha información, el proveedor de servicios tiene que conocer los recursos disponibles a través de los proveedores de datos<sup>7</sup>.

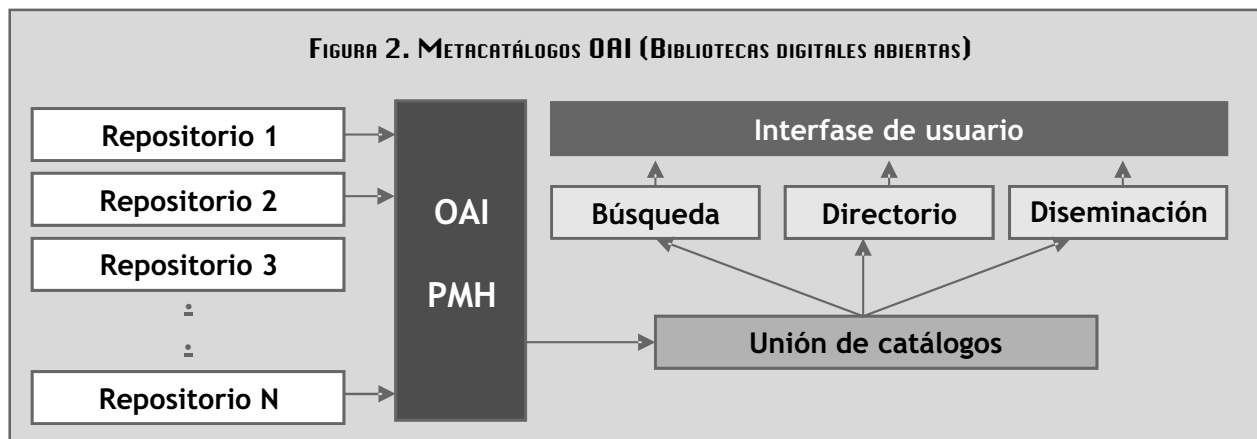
---

7 (Un directorio de proveedores de datos existentes y registrados, se puede obtener de: (<http://www.openarchives.org/Register/BrowseSites>), y una lista por país la puede ver en: (<http://archives.eprints.org/index.php?action=browse>), luego debe revisar que el proveedor de datos este accesible para que el proveedor de servicios pueda capturar sus metadatos. Esta tarea de recolección se realiza generalmente de forma automática siguiendo un proceso en lotes (Batch), que se realiza en forma periódica (usualmente cada 24 horas).



Uno de los casos más importantes de aplicación es la Biblioteca Digital Interconectada de Tesis y Dissertaciones (NDLTD), la cual contiene colecciones

de múltiples e independientes documentos electrónicos, conformada por distintas universidades (Suleman *et al*, 2001).



## RECOLECCIÓN FRENTE A FEDERACIÓN

Adicionalmente el modelo UPS, manifiesta dos métodos de trabajo que debe contemplar el proveedor de servicios para lograr la interoperabilidad de acuerdo a la ubicación de la información y la relación de trabajo de los repositorios, sabiendo de antemano que la información se encuentra ampliamente distribuida:

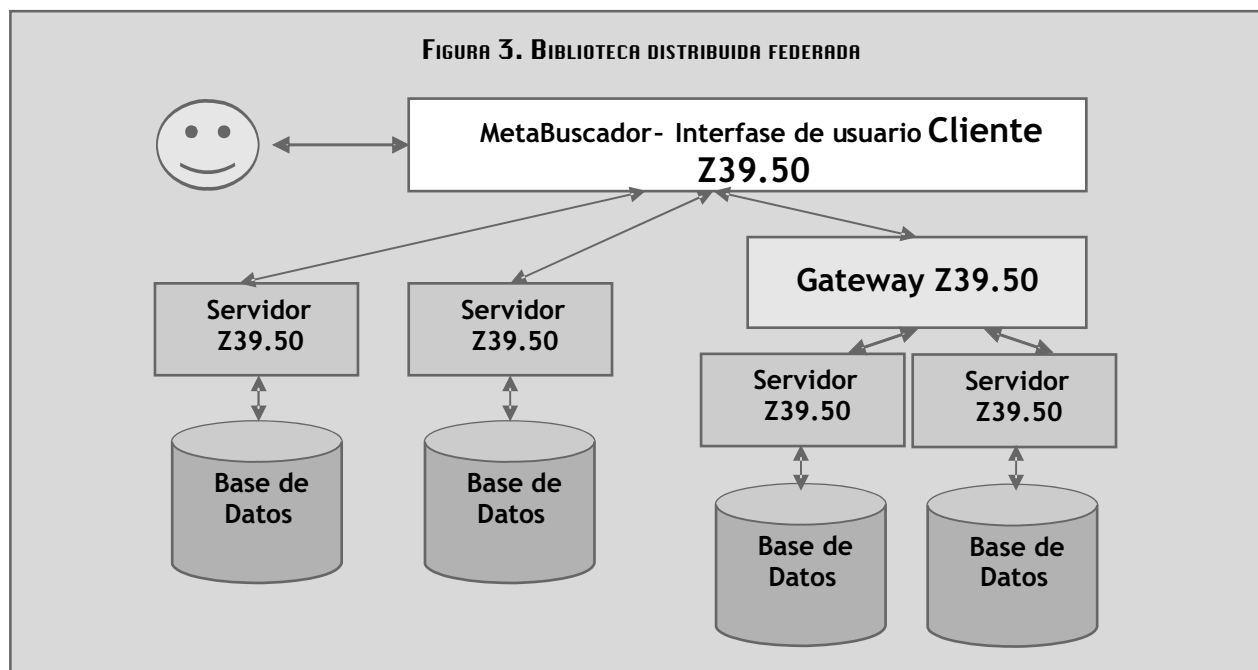
- ◆ **Federación:** bajo este enfoque la información y los servicios están almacenados descentralizadamente. El proveedor de servicios, cada vez que el usuario desea realizar alguna búsqueda de información, ejecuta los servicios remotamente de acuerdo a las necesidades del usuario y la lista de proveedores de datos que contenga, por ejemplo metabúsqueda distribuida y federada con Z39.50.

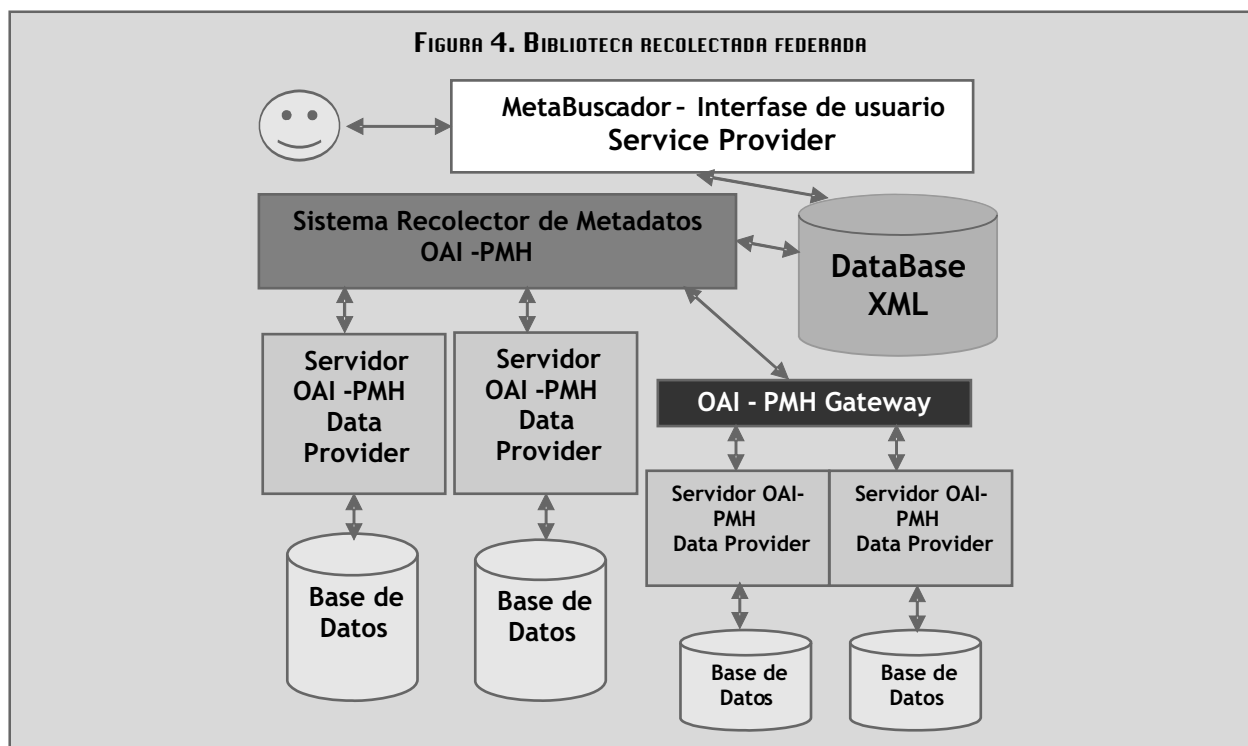
- ◆ **Recolección:** toda la información que poseen los proveedores de datos (datos/metadatos) se transfieren continuamente al proveedor de servicios, donde se almacenan en forma centralizada, y desde allí los usuarios realizan sus consultas utilizando los servicios ofrecidos por el sistema central. Por ejemplo, biblioteca digital distribuida y federada.

El modelo de Federación requiere más esfuerzo en cada fuente, pero es más fácil para el sistema local, ya que no hay que hacer procesamiento, sino hacer conexión con cada repositorio, formularle una estrategia de búsqueda y mostrar resultados finales al usuario. Para que este modelo sea eficiente los tiempos individuales de procesamiento y envío de la información desde la fuente al sistema cen-

tral deben ser muy cortos, ya que una demora considerable afectaría la entrega de resultados y ocasionaría una demora general del sistema.

Recolectar requiere más esfuerzo en el sistema local, en el diseño de los servicios y la recolección continua y automatizadas de la información, pero la implementación para los proveedores de datos es más sencilla. Adicionalmente, al estar la información centralizada en el mismo servidor, la velocidad en obtener las respuestas a las consultas de los usuarios, dependerían exclusivamente del poder de procesamiento local, independientemente de la información almacenada. La actual tendencia de la iniciativa de archivos abiertos se basa en recolección, a través de la implantación del protocolo de recolección de metadatos **OAI-PMH** tanto en proveedores de datos como de servicios.





### PROTOCOLO DE RECOLECCIÓN DE METADATOS (OAI-PMH)

El protocolo de recolección de metadatos (OAI-PMH), es el mecanismo de trabajo que debe ser implementado en proveedores de datos y de servicios para poder comunicarse entre sí (interoperabilidad). Este protocolo define todos los «comandos» o «verbos» necesarios para poder recolectar metadatos de forma automática y normalizada. Este protocolo define en solo seis (6) verbos básicos, y algunos parámetros complementarios un mecanismo para transferir metadatos, inicialmente se diseñó para ser implementado lo más sen-

cillo posible y, pensando en la gran disparidad en los niveles y forma de catalogación de los documentos, se optó por utilizar el formato de metadatos estándar más reducido posible: *Dublin Core*<sup>8</sup> sin calificadores, aunque se dejó abierto el mecanismo de utilizar cualquier formato de metadatos, estándar o de creación propia.

Este formato *Dublin Core reducido*, es el único modelo de metadatos que es obligatorio soportar en los proveedores de datos y servicios, y se debe poder invocar bajo el nombre (*oai\_dc*), y el esquema de validación asociado en XML se encuentra, en la página principal de la iniciativa:

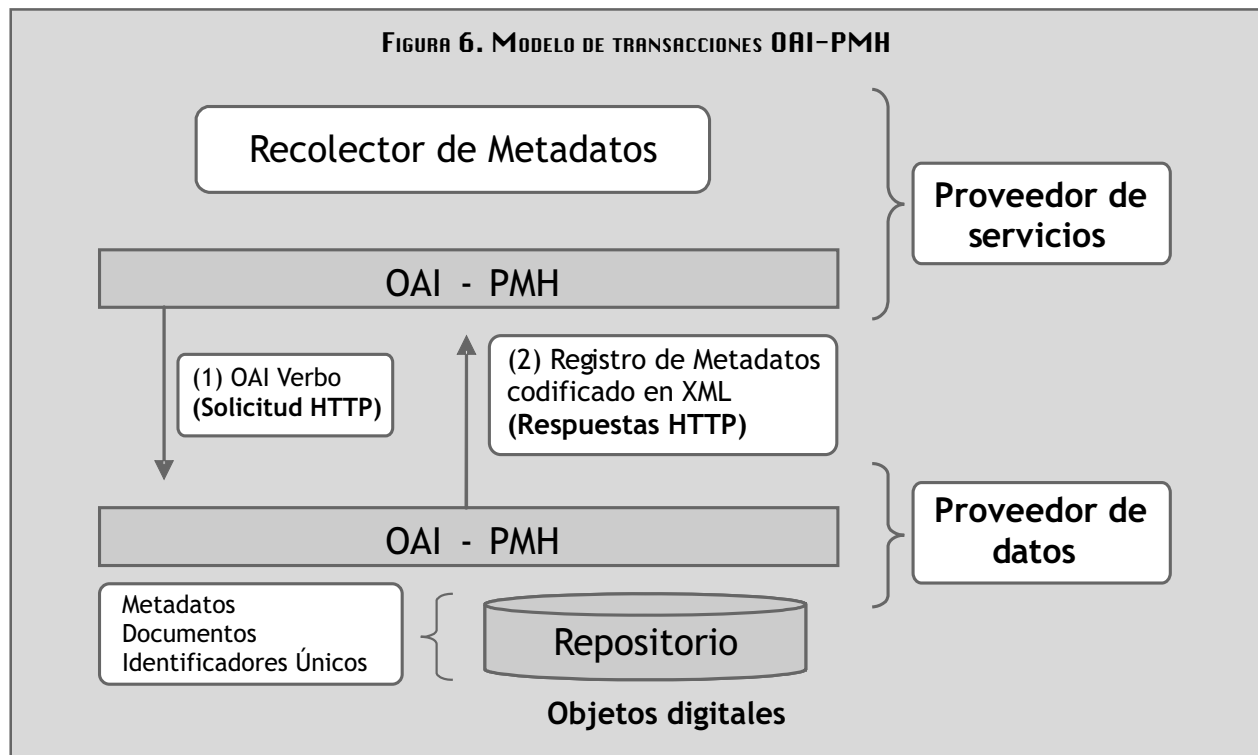


8 <http://www.dublincore.org/>

En la figura anterior, se muestran las características del modelo de metadatos oai\_dc, cuando nos referimos a espacios de nombre (*Namespace*) en XML, nos referimos a una manera de especificación que sirve para evitar las colisiones entre elementos del mismo nombre, y en general, para distinguir los distintos grupos de elementos en un mismo documento, por ejemplo distinguir claramente los elementos del modelo *Dublin Core* reducido respecto al modelo *Dublin Core* completo con calificadores, ya que ambos poseen etiquetas comunes y que se invocan igual (Título - *Title*). Cada espacio de nombres se asocia con una URL, que sólo sirve como

identificador único y no tiene por qué ser una dirección válida (Arranz, 2005). El esquema (*Schema*) es un tipo de documento en XML, que permite hacer restricciones más precisas sobre los documentos XML asociados (tanto en estructura como en tipos de datos), en este caso el esquema verifica que el modelo de metadatos utilizado (*Dublin Core reducido*) esté bien diseñado y cumpla todas las normas establecidas; por lo tanto, se busca que la información compartida corresponda a un documento válido.

## ¿CÓMO TRABAJA OAI-PMH?



La figura anterior (Cole *et al*, 2003) muestra el funcionamiento del protocolo OAI-PMH, básicamente éste utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un proveedor de servicios y un proveedor de datos. Estas transacciones se realizan utilizando los métodos GET o POST, que representan las operaciones de recepción y envío de información entre un programa

cliente servidor a través del protocolo HTTP, éstas se han utilizado comúnmente en la creación de programas CGI (pasarela de interfase común) y constan de:

- ◆ Una dirección URL de un archivo que ejecuta un procedimiento.

- ◆ Una lista de parámetros con la forma de pares del tipo: nombre\_variable = valor.

Un ejemplo de una transacción HTTP es:  
(<http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=Identify>)

Donde <http://biblioteca.unisabana.edu.co/tesis/oai.php> indica la ubicación del archivo **oai.php** en el servidor de Internet en la Biblioteca de la Universidad de La Sabana, este archivo cuando es invocado desde un navegador de Internet, ejecuta unos

procedimientos internos que son condicionados generalmente a unos parámetros asociados, el parámetro en este ejemplo es la palabra «*verb*», cuyo valor es «*Identify*»; es conveniente observar que el valor de los parámetros distingue la diferencia entre mayúsculas y minúsculas. OAI-PMH establece que siempre debe existir en cualquier llamado de transacción el parámetro «*verb*», el cual indica al proveedor de datos, el tipo de procedimiento requerido por el proveedor de servicios, en caso de omitirse este parámetro el sistema nos enviará el siguiente mensaje:

**FIGURA 7. RESPUESTA DE ERROR – PROVEEDOR DE DATOS**

```
<?xml version="1.0" encoding="UTF-8" ?>

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://
www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://
www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">

  <responseDate>2005-06-22T18:59:11Z</responseDate>
  <request>http://biblioteca.unisabana.edu.co/tesis/oai.php</request>
  <error code="badVerb">Verbo utilizado no existe en OAI</error>
</OAI-PMH>
```

## ¿CÓMO SON LAS RESPUESTAS?

Las respuestas que proporciona el proveedor de datos, al ser invocada una transacción HTTP, son presentadas en texto XML, aunque existe la posibilidad de responder con un archivo comprimido que contiene el documento XML antes mencionado. El formato de respuesta contiene las siguientes partes (Muller, 2002):

1. Declaración de documento XML:

(`<?xml version="1.0" encoding="UTF-8" ?>`)

2. Espacios de nombre, que hacen invocación a los elementos constitutivos de un documento OAI, con sus esquemas de validación asociados:

(xmlns, xmlns:xsi, xsi:schemaLocation)

3. Tres elementos hijos:

- ◆ **responseDate** (UTC datetime): corresponde a la fecha y hora en formato UTC, en la cual el proveedor de datos atendió la transacción HTTP:

◆ (`<responseDate>2005-06-24T20:40:35Z</responseDate>`)

**REQUEST. CORRESPONDE A LA REFERENCIA QUE INDICA EL VERBO JUNTO CON TODOS LOS PARÁMETROS SOLICITADOS**

(`<request verb="ListRecords" metadata Prefix="oai_dc">`

from=>2004 01 03">http://biblioteca.unisabana.edu.co/tesis/oai.php</request>)

metadatos, información del repositorio, etc).

- ◆ Esta parte corresponde al cuerpo de la respuesta dada por el proveedor de datos y puede ser de dos tipos:

- b) **Error:** en caso de obtener algún error en la transacción, se genera un informe específico acerca del problema.

- a) **Datos solicitados:** de acuerdo al requerimiento expresado en el verbo (Tipos,

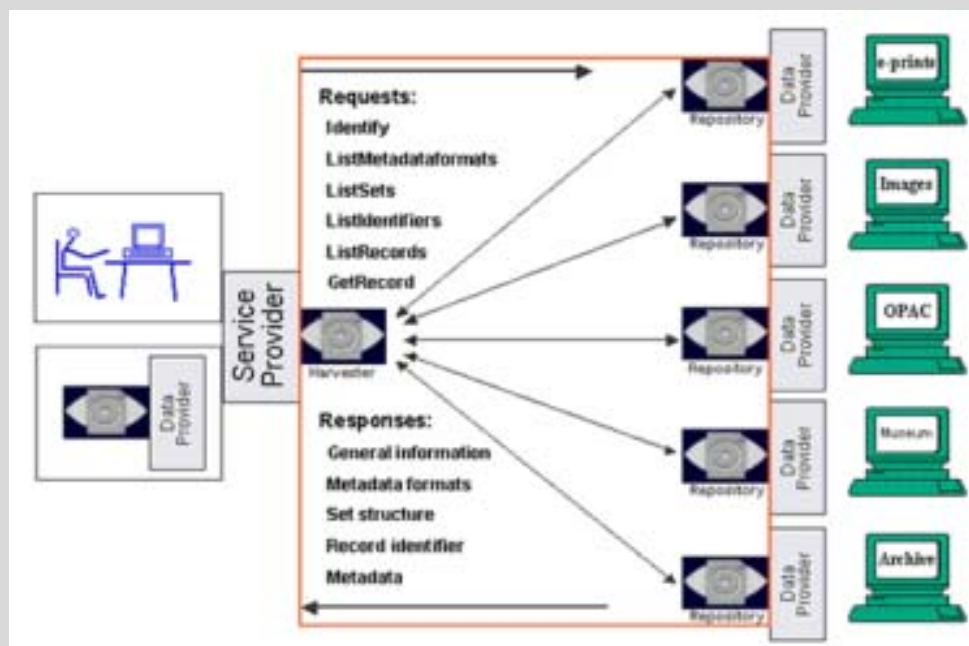
Existen seis verbos o procedimientos definidos en OAI-PMH y los cuales se pueden invocar a un proveedor de datos:

**FIGURA 8. LISTA DE VERBOS OAI-PMH**

Verbo	Función
<b>Metadatos acerca del repositorio</b>	
<i>Identify</i>	Descripción del repositorio
<i>ListMetadataFormats</i>	Lista de todos los tipos de metadatos soportados en el repositorio
<i>ListSets</i>	Lista de las particiones lógicas de información en el repositorio
<b>Verbos de recolección de metadatos</b>	
<i>ListIdentifiers</i>	Lista de los identificadores OAI únicos existentes en el repositorio
<i>ListRecords</i>	Lista de n registros de metadatos
<i>GetRecord</i>	Lista de un solo registro de metadatos

Fuente: Lagoze *et al.* 2002

**FIGURA 9. MODELO DE RECOLECCIÓN OAI-PMH**





<b>Verbo</b>	<b>Identify</b>
<b>Descripción</b>	<p>Recupera información acerca del proveedor de datos, sus colecciones y sus políticas, este informa acerca de:</p> <ul style="list-style-type: none"> <li>• El nombre del proveedor de datos</li> <li>• Las versiones de OAI soportadas</li> <li>• La dirección del administrador responsable</li> <li>• La fecha del registro de metadatos más antiguo</li> <li>• Opciones de compresión, edición y granularidad del sistema</li> <li>• Un ejemplo acerca del manejo de identificadores en el sistema</li> <li>• Proveedores de datos conocidos (Friends), para futuras recolecciones</li> </ul>
<b>Parámetros</b>	Ninguno
<b>Modelo</b>	<a href="http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=Identify">http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=Identify</a>
<b>Ejemplo</b>	<pre> &lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/     http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"&gt;   &lt;responseDate&gt;2005-06-21T22:49:06Z&lt;/responseDate&gt;   &lt;request verb="Identify"&gt;http://biblioteca.unisabana.edu.co/tesis/oai.php&lt;/request&gt;    &lt;Identify&gt;     &lt;repositoryName&gt;Biblioteca Universidad De La Sabana - Catálogo de Tesis y     Disertaciones&lt;/repositoryName&gt;     &lt;baseURL&gt;http://biblioteca.unisabana.edu.co/tesis/oai.php&lt;/baseURL&gt;     &lt;protocolVersion&gt;2.0&lt;/protocolVersion&gt;     &lt;adminEmail&gt;felipe.gomez3@unisabana.edu.co&lt;/adminEmail&gt;     &lt;earliestDatestamp&gt;2003-12-09&lt;/earliestDatestamp&gt;     &lt;deletedRecord&gt;no&lt;/deletedRecord&gt;     &lt;granularity&gt;YYYY-MM-DD&lt;/granularity&gt;     &lt;compression&gt;gzip&lt;/compression&gt;     &lt;compression&gt;deflate&lt;/compression&gt;      &lt;description&gt;     &lt;oai-identifier xmlns=http://www.openarchives.org/OAI/2.0/oai-identifier       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"       xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier         http://www.openarchives.org/OAI/2.0/oai-identifier.xsd"&gt;       &lt;scheme&gt;oai&lt;/scheme&gt;       &lt;repositoryIdentifier&gt;biblioteca.unisabana.edu.co&lt;/repositoryIdentifier&gt;       &lt;delimiter&gt;:&lt;/delimiter&gt;       &lt;sampleIdentifier&gt; oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001&lt;/sampleIdentifier&gt;     &lt;/oai-identifier&gt;     &lt;/description&gt;      &lt;description&gt;     &lt;friends xmlns="http://www.openarchives.org/OAI/2.0/friends/"       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"       xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/         http://www.openarchives.org/OAI/2.0/friends.xsd"&gt;       &lt;baseURL&gt;http://www.scielo.org/oai.php&lt;/baseURL&gt;       &lt;baseURL&gt;http://dspace.uniandes.edu.co/oai.php&lt;/baseURL&gt;     &lt;/friends&gt;     &lt;/description&gt;   &lt;/Identify&gt; &lt;/OAI-PMH&gt; </pre>

<b>Verbo</b>	<b>GetRecord</b>
<b>Descripción</b>	Este verbo retorna los metadatos de un registro solicitado según el formato especificado (por defecto oai_dc)
<b>Parámetros</b>	<p>Requiere dos parámetros obligatorios:</p> <ul style="list-style-type: none"> <li>• <b>Identifier</b>: Identificador único del documento en formato URN (Lagoze et al, 2002)</li> <li>• <b>metadataPrefix</b>: Modelo de metadatos a desplegar, como mínimo debe especificar <b>oai_dc</b>, para conocer los metadatos soportador por el proveedor de datos, utilice el verbo “<b>ListMetadataFormats</b>”</li> </ul>
<b>Modelo</b>	<a href="http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=GetRecord&amp;identifier=oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001&amp;metadataPrefix=oai_dc">http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=GetRecord&amp;identifier=oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001&amp;metadataPrefix=oai_dc</a>
<b>Ejemplo</b>	<pre>&lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/     http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"&gt;   &lt;responseDate&gt;2005-06-22T19:14:18Z&lt;/responseDate&gt;   &lt;request verb="GetRecord" identifier="oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001"     metadataPrefix="oai_dc"&gt;http://biblioteca.unisabana.edu.co/tesis/oai.php&lt;/request&gt;    &lt;GetRecord&gt;   &lt;record&gt;   &lt;header&gt;     &lt;identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001&lt;/identifier&gt;     &lt;timestamp&gt;2004-05-07&lt;/timestamp&gt;     &lt;setSpec&gt;Unisabana&lt;/setSpec&gt; &lt;setSpec&gt;Tesis&lt;/setSpec&gt;&lt;setSpec&gt;Tesis:Pregrado&lt;/setSpec&gt;   &lt;/header&gt;   &lt;metadata&gt;     &lt;oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"       xmlns:dc="http://purl.org/dc/elements/1.1/"       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"       xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/         http://www.openarchives.org/OAI/2.0/oai_dc.xsd"&gt;       &lt;dc:title&gt;Plan de mercadeo de Cemento &lt;/dc:title&gt;       &lt;dc:creator&gt;Santacoloma Jaramillo, Alejandro&lt;/dc:creator&gt;       &lt;dc:contributor&gt;Correa Correa, Luis Fernando ases.&lt;/dc:contributor&gt;       &lt;dc:subject&gt;ESTUDIOS DE MERCADEO&lt;/dc:subject&gt;       &lt;dc:subject&gt;INDUSTRIA DEL CEMENTO&lt;/dc:subject&gt;       &lt;dc:publisher&gt;Universidad de La Sabana - Biblioteca, 2000.&lt;/dc:publisher&gt;       &lt;dc:date&gt;2000-08-10&lt;/dc:date&gt;       &lt;dc:language&gt;spa&lt;/dc:language&gt;       &lt;dc:coverage&gt;Chía, Cundinamarca&lt;/dc:coverage&gt;       &lt;dc:format&gt;application/pdf&lt;/dc:format&gt;       &lt;dc:type&gt;TESIS&lt;/dc:type&gt;       &lt;dc:identifier&gt;http://biblioteca.unisabana.edu.co/catalogo/openurl.php?base=dublin&amp;mfn=000023       &lt;/dc:identifier&gt;&lt;dc:identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis-221.2000-001&lt;/dc:identifier&gt;       &lt;dc:description&gt;Requerimientos del sistema : Acrobat Reader. -         Incluye bibliografía y anexos&lt;/dc:description&gt;       &lt;dc:description&gt;Administrador de Empresas, Facultad de Ciencias Económicas y Administrativas.         Programa de Administración de Empresas, 2000.&lt;/dc:description&gt;       &lt;dc:rights&gt;Uso General- Ver: http://biblioteca.unisabana.edu.co/tesis/general.htm&lt;/dc:rights&gt;        &lt;dc:source&gt;http://biblioteca.unisabana.edu.co/catalogo/inc/openurl.php?mfn=000023&amp;base=Dublín       &lt;/dc:source&gt;     &lt;/oai_dc:dc&gt;   &lt;/metadata&gt;   &lt;/record&gt; &lt;/GetRecord&gt; &lt;/OAI-PMH&gt;</pre>

Verbo	<b>ListMetadataFormats</b>
Descripción	Muestra la lista de formatos de metadatos disponibles en la colección (estandarizados o propios)
Parámetros	<p>Posee un parámetro opcional:</p> <ul style="list-style-type: none"> <li><b>Identifier:</b> Identificador único del documento en formato URN, al cual se pueden asociar uno o más formatos de metadatos específicos.</li> </ul>
Modelo	http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListMetadataFormats
	http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListMetadataFormats&identifier=oai:biblioteca.unisabana.edu.co:tesis-221.2001-003
Ejemplo	<pre>&lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"&gt;   &lt;responseDate&gt;2005-06-22T20:20:16Z&lt;/responseDate&gt;   &lt;request verb="ListMetadataFormats" identifier="oai:biblioteca.unisabana.edu.co:Tesis-221.2001-003"&gt;http://biblioteca.unisabana.edu.co/tesis/oai.php&lt;/request&gt;   &lt;ListMetadataFormats&gt;    &lt;metadataFormat&gt;     &lt;metadataPrefix&gt;oai_dc&lt;/metadataPrefix&gt;     &lt;schema&gt;http://www.openarchives.org/OAI/2.0/oai_dc.xsd&lt;/schema&gt;     &lt;metadataNamespace&gt;http://www.openarchives.org/OAI/2.0/oai_dc/   &lt;/metadataNamespace&gt;   &lt;/metadataFormat&gt;    &lt;metadataFormat&gt;     &lt;metadataPrefix&gt;mtd_col&lt;/metadataPrefix&gt;     &lt;schema&gt;http://biblioteca.unisabana.edu.co/tesis/schemas/mtd_col.xsd&lt;/schema&gt;     &lt;metadataNamespace&gt;http://biblioteca.unisabana.edu.co/tesis/schemas/&lt;/metadataNamespace&gt;   &lt;/metadataFormat&gt;    &lt;metadataFormat&gt;     &lt;metadataPrefix&gt;marcxml&lt;/metadataPrefix&gt;     &lt;schema&gt;http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd&lt;/schema&gt;</pre>

Verbo	ListSets
Descripción	Especifica los conjuntos o particiones lógicas en que está estructurada la colección, esta puede dividirse según criterio del administrador del proveedor de datos siguiendo la notación URI (rfc 2396) (Berners-Lee, 1998), por ejemplo, se puede dividir la colección en subcolecciones temáticas, temporales, por tipo de colección o por fuente de información (creador de los datos).
Parámetros	<p>Posee un parámetro opcional:</p> <ul style="list-style-type: none"> <li><b>resumptionToken:</b> cuando la lista de particiones lógicas del sistema es demasiado grande (generalmente más de 100), se utiliza este parámetro, para desplegar los resultados en varios llamados de transacciones HTTP.</li> </ul>
Modelo	<p><a href="http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListSets">http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListSets</a></p> <p><a href="http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListSets&amp;resumptionToken=1">http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListSets&amp;resumptionToken=1</a></p>
Ejemplo	<pre>&lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/     http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"&gt;   &lt;responseDate&gt;2005-06-22T20:31:08Z&lt;/responseDate&gt;   &lt;request verb="ListSets"&gt;http://biblioteca.unisabana.edu.co/tesis/oai.php&lt;/request&gt;    &lt;ListSets&gt;   &lt;set&gt;     &lt;setSpec&gt;Unisabana&lt;/setSpec&gt;     &lt;setName&gt;Registros de la Universidad de La Sabana&lt;/setName&gt;     &lt;setDescription&gt;       &lt;oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"         xmlns:dc="http://purl.org/dc/elements/1.1/"         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"         xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/           http://www.openarchives.org/OAI/2.0/oai_dc.xsd"&gt;         &lt;dc:description&gt;Todos los registros de la U. de La Sabana&lt;/dc:description&gt;         &lt;dc:title&gt;Titulo nuevo&lt;/dc:title&gt;       &lt;/oai_dc:dc&gt;     &lt;/setDescription&gt;   &lt;/set&gt;    &lt;set&gt;     &lt;setSpec&gt;Tesis&lt;/setSpec&gt;     &lt;setName&gt;Tesis de la Univerisdad de la Sabana&lt;/setName&gt;     &lt;setDescription&gt;       &lt;oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"         xmlns:dc="http://purl.org/dc/elements/1.1/"         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"         xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/           http://www.openarchives.org/OAI/2.0/oai_dc.xsd"&gt;         &lt;dc:description&gt;Todas las tesis digitales&lt;/dc:description&gt;       &lt;/oai_dc:dc&gt;     &lt;/setDescription&gt;   &lt;/set&gt;    &lt;/ListSets&gt; &lt;/OAI-PMH&gt;</pre>

<b>Verbo</b>	<b>ListIdentifiers</b>
<b>Descripción</b>	Muestra la lista de identificadores existentes en la colección
<b>Parámetros</b>	<p>Posee tres parámetros opcionales:</p> <ul style="list-style-type: none"> <li>• <b>set:</b> determina el nombre de la partición lógica sobre la que se desea trabajar, lo que me permite recolectar resultados parciales de acuerdo a los intereses del proveedor de servicios.</li> <li>• <b>from:</b> determina la fecha inicial en formato normalizado<sup>9</sup> de los metadatos que se desean recuperar.</li> <li>• <b>until:</b> determina la fecha final, de los metadatos a recuperar. Tanto la fecha inicial como la fecha final, pueden aparecer una o ambas de forma indistinta para el proveedor de datos.</li> </ul> <p>Posee un parámetro exclusivo</p> <ul style="list-style-type: none"> <li>• <b>resumptionToken:</b> permite desplegar los resultados en varios llamados de transacciones HTTP, se utiliza para el control de flujo.</li> </ul>
<b>Modelo</b>	<p><code>http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListIdentifiers&amp;metadataPrefix=oai_dc</code></p> <p><code>http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListIdentifiers&amp;metadataPrefix=oai_dc&amp;from=2002-12-02</code></p> <p><code>http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListIdentifiers&amp;resumptionToken=1</code></p>
<b>Ejemplo</b>	<pre>&lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI PMH.xsd"&gt;   &lt;responseDate&gt;2005 06 23T22:35:32Z&lt;/responseDate&gt;   &lt;request verb="ListIdentifiers" metadataPrefix="oai_dc"&gt;     http://biblioteca.unisabana.edu.co/tesis/oai.php   &lt;/request&gt;    &lt;ListIdentifiers&gt;     &lt;header&gt;       &lt;identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis 386.2002 010&lt;/identifier&gt;       &lt;timestamp&gt;2005 04 16&lt;/timestamp&gt;       &lt;setSpec&gt;Unisabana&lt;/setSpec&gt;       &lt;setSpec&gt;Tesis&lt;/setSpec&gt;       &lt;setSpec&gt;Tesis:Pregrado&lt;/setSpec&gt;     &lt;/header&gt;     ...     &lt;header&gt;       &lt;identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis 251.2000 017&lt;/identifier&gt;       &lt;timestamp&gt;2003 12 09&lt;/timestamp&gt;       &lt;setSpec&gt;Unisabana&lt;/setSpec&gt;       &lt;setSpec&gt;Tesis&lt;/setSpec&gt;       &lt;setSpec&gt;Tesis:Pregrado&lt;/setSpec&gt;     &lt;/header&gt;      &lt;resumptionToken expirationDate="2005 06 23T23:35:34Z" completeListSize="2364" cursor="0"&gt;       1258     &lt;/resumptionToken&gt;   &lt;/ListIdentifiers&gt; &lt;/OAI PMH&gt;</pre>

<sup>9</sup> Todas los formatos de fecha y tiempo, deben estar codificados uniformemente utilizando el estándar ISO8601, y estas deben ser expresadas con la notación UTC, tal como indica el protocolo OAI-PHM (<http://www.w3.org/TR/NOTE-datetime>)

Verbo	ListRecords
Descripción	Muestra los metadatos de los registros solicitados, en el formato de metadatos estipulado (básicamente oai_dc)
Parámetros	Posee los mismos parámetros del verbo "ListIdentifiers"
Modelo	<p>http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListRecords&amp;metadataPrefix=oai_dc&amp;from=2004-01-03&amp;set=tesis:pregrado</p> <p>http://biblioteca.unisabana.edu.co/tesis/oai.php?verb=ListRecords&amp;resumptionToken=56</p>
Ejemplo	<pre>&lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;OAI PMH xmlns="http://www.openarchives.org/OAI/2.0/"   xmlns:xsi="http://www.w3.org/2001/XMLSchema instance"   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/     http://www.openarchives.org/OAI/2.0/OAI PMH.xsd"&gt;   &lt;responseDate&gt;2005 06 23T22:49:56Z&lt;/responseDate&gt;   &lt;request verb="ListRecords" metadataPrefix="oai_dc" from="2004 01 03"&gt;     http://biblioteca.unisabana.edu.co/tesis/oai.php   &lt;/request&gt;    &lt;ListRecords&gt;   &lt;record&gt;   &lt;header&gt;   &lt;identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis 386.2002 010&lt;/identifier&gt;   &lt;timestamp&gt;2005 04 16&lt;/timestamp&gt;   &lt;setSpec&gt;Unisabana&lt;/setSpec&gt;   &lt;setSpec&gt;Tesis&lt;/setSpec&gt;   &lt;setSpec&gt;Tesis:Pregrado&lt;/setSpec&gt;   &lt;/header&gt;   &lt;metadata&gt;   &lt;oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"     xmlns:dc="http://purl.org/dc/elements/1.1/"     xmlns:xsi="http://www.w3.org/2001/XMLSchema instance"     xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/       http://www.openarchives.org/OAI/2.0/oai_dc.xsd"&gt;     &lt;dc:title&gt;Desarrollo de la competencia semántica a través de la música en niños de básica&lt;/dc:title&gt;     &lt;dc:creator&gt;Fernández Riaño, Andrea&lt;/dc:creator&gt;     &lt;dc:contributor&gt;Martínez Lopera Blanca Elena , ases.&lt;/dc:contributor&gt;     &lt;dc:subject&gt;SEMANTICA&lt;/dc:subject&gt;     &lt;dc:subject&gt;MUSICOTERAPIA&lt;/dc:subject&gt;     &lt;dc:publisher&gt;Universidad de La Sabana Biblioteca, 2002.&lt;/dc:publisher&gt;     &lt;dc:date&gt;2002 11 18&lt;/dc:date&gt;     &lt;dc:language&gt;spa&lt;/dc:language&gt;     &lt;dc:coverage&gt;Chía, Cundinamarca&lt;/dc:coverage&gt;     &lt;dc:format&gt;application/pdf&lt;/dc:format&gt;     &lt;dc:type&gt;TESIS&lt;/dc:type&gt;     &lt;dc:identifier&gt;386.2002 010&lt;/dc:identifier&gt;     &lt;dc:identifier&gt;oai:biblioteca.unisabana.edu.co:Tesis 386.2002 010&lt;/dc:identifier&gt;     &lt;dc:description&gt;Requerimientos del sistema : Windows 95 ó superior&lt;/dc:description&gt;     &lt;dc:description&gt;Licenciado en Lingüística y Literatura, Facultad de Educación. Departamento de Lingüística       y Literatura, 2002.&lt;/dc:description&gt;     &lt;dc:rights&gt;Uso General Ver: http://biblioteca.unisabana.edu.co/tesis/general.htm&lt;/dc:rights&gt;     &lt;dc:source&gt;http://biblioteca.unisabana.edu.co/catalogo/inc/openurl.php?mf=000001&amp;base=Dublin   &lt;/dc:source&gt;   &lt;/oai_dc:dc&gt;   &lt;/metadata&gt;   &lt;/record&gt;   &lt;/ListRecords&gt; &lt;/OAI PMH&gt;</pre>

## CARACTERÍSTICAS DE LOS REGISTROS DE METADATOS

Los registros que devuelve el proveedor de datos cuando se invocan los verbos *ListRecords* y *GetRecord*, deben estar contruidos según especifica el protocolo en tres partes básicas (cabecera, metadatos asociados y anotaciones especiales «*about*»), cada uno de estos componentes es descrito a continuación:

- ◆ **header** (obligatorio): cabecera que antepone a los metadatos información general de este más no datos bibliográficos.
- ◆ **identifier** (obligatorio: solo 1): contiene el identificador único del registro en el repositorio e Internet en notación URN.

(<identifier>oai:biblioteca.unisabana.edu.co:Tesis-386.2002-010</identifier>)

- ◆ **datestamp** (obligatorio: solo 1): corresponde a la última fecha en que fue modificado / creado el registro, en formato UTC.

(<datestamp>2005 04 16</datestamp>)

- ◆ **setSpec** elements (opcional: 0, 1 o más): establece a que particiones lógicas está asociado este registro:

(<setSpec>Unisabana</setSpec>  
<setSpec>Tesis</setSpec>  
<setSpec>Tesis:Pregrado</setSpec>)

- ◆ **Status** (opcional: 0 o 1): determina si el registro fue eliminado del sistema.
- ◆ **metadata** (obligatorio): contiene los metadatos asociados (datos bibliográficos) a un documento

- ◆ **Metadatos en XML**: son los metadatos desplegados en XML, bajo un formato específico (por defecto oai\_dc).

- ◆ **about** (opcional): características adicionales al registro:

- ◆ **Copyright** (Derechos de autor): contiene información adicional del autor y los derechos asociados al documento.

- ◆ **Provenance** (Opcional): si el registro ha sido recolectado o tomado de otro proveedor de datos, aquí aparece información de éste, con el fin de evitar duplicación de información.

## ¿CÓMO CREAR UN ARCHIVO ABIERTO?

Cualquier persona con algo de conocimientos en Internet puede implementar un proveedor de datos o un proveedor de servicios y realizar una óptima implementación del protocolo OAI-PMH, esto con el fin de poner a disposición de la comunidad los datos que hasta ahora estaban escondidos en nuestras bases de datos o repositorios institucionales.

## PROVEEDOR DE DATOS

Para construir un proveedor de datos, se puede iniciar desde la metodología más sencilla que corresponde en crear un archivo de texto con la información que poseemos y almacenarla siguiendo la estructura de un archivo XML, y luego ponemos este archivo creado en un servidor que se pueda visualizar en Internet. Este archivo debe estar formateado siguiendo la estructura propuesta por la iniciativa de archivos abiertos para repositorios estáticos (Van de Sompel, 2002).

La segunda forma consiste en adaptar nuestros repositorios de documentos y bibliotecas digitales

existentes, de tal forma que permitiera recibir las peticiones OAI-PMH emitidas por un proveedor de servicios, éste interrogará nuestro sistema de información o base de datos y pudiera obtener una respuesta de acuerdo al protocolo. En este sentido bastaría con disponer de un servidor de Internet y desarrollar una aplicación (por ejemplo en Perl, PHP, ASP o JSP) que funcionara en entorno *Web*. Para hacer esto debemos conocer detalladamente nuestro sistema y así realizar una integración óptima; como ventaja podemos encontrar mucho código fuente desarrollado en los principales lenguajes de programación de Internet listo para adaptar y utilizar. Si no poseemos suficientes conocimientos en programación podemos conseguir libremente en Internet varias aplicaciones *middleware* genéricas que, sin tener mayores complicaciones, se pueden adaptar a nuestros sistemas mediante una parametrización en la cual brindamos datos de acceso y establecemos la equivalencia de los campos de nuestra base con los campos básicos asociados al estándar *Dublin Core* (*oai\_dc*), como ejemplos de estos programas prediseñados podemos citar<sup>6</sup>:

- ◆ **ASP OAI 2.0 Data Provider:** *The University of Illinois at Urbana-Champaign, Open Archives Initiative Metadata Harvesting Project* (<http://oai.grainger.uiuc.edu/>) (UIUC)
- ◆ **ibiblio's PHP:** implementación de un OAI-PMH proveedor de datos.
- ◆ **OAICat:** sistema Base de implementación de OCLC.
- ◆ **vOAI:** generador de servidores OAI, Universidad de las Américas, Puebla.

- ◆ **MOD\_OAI:** un programa para el servidor de Internet Apache, desarrollado por *Old Dominion University* y la *Andrew W. Mellon Foundation*.

La tercera manera de construir un proveedor de datos, consiste en descargar e instalar un programa de la gran cantidad de alternativas de *software* existente en Internet que soportan el protocolo de OAI, y que bajo la premisa del movimiento del acceso abierto (Open Access) se consiguen libremente, existe un directorio oficial de *software* registrado (<http://www.openarchives.org/tools/tools.html>) que han construido distintas organizaciones y universidades cuyo objetivo es el de poner a disposición del público toda la información académica y científica que se produce y que actualmente no es posible consultar fácilmente. Con estos programas cualquier institución (universidad o centro de investigación) puede crear su propio archivo documental al tiempo que son compatibles con OAI-PMH, ejemplos de algunos de los programas más utilizados son: Dspace, Eprints y CDSWare.

## PROVEEDOR DE SERVICIOS

Un sistema proveedor de servicios, puede verse desde un *software* que sirva como recolector de metadatos y permita consultar la información contenida en éstos, hasta sistemas de bibliotecas digitales complejos, que involucren una gran cantidad de servicios digitales, los principales programas de uso libre son:

- ◆ **ARC harvester and search engine** (<http://sourceforge.net/projects/oaiarc/>)

6 Una excelente evaluación y comparación detallada de los principales programas existentes (proveedores de datos y servicios) que son de uso libre (free open source) (<http://www.library.uq.edu.au/escholarship/software.pdf>)



- ◆ **PKP OAI Harvester:** (<http://www.pkp.ubc.ca/pkp-harvester/>)
- ◆ **My.OAI:** (<http://www.myoai.com/downloads/>)
- ◆ **Celestial:** (<http://celestial.eprints.org/>)

Una vez puesto en funcionamiento nuestro proveedor de datos o de servicios, es conveniente registrarlo en la página de la iniciativa (<http://www.openarchives.org/community/>). Allí el sistema realiza las pruebas de implantación a nuestro servicio y nos publica en el directorio oficial correspondiente.

## CONCLUSIONES

Cada vez que busco en Internet, proyectos de repositorios institucionales y bibliotecas digitales, veo que éstos ya se encuentran trabajando la iniciativa de archivos abiertos, porque buscan desde sus inicios el obtener una mayor visibilidad en Internet y lograr la mayor interoperabilidad con colecciones afines. En este momento, cuando hablamos de bibliotecas digitales en el ámbito universitario, tenemos que pensar indudablemente en OAI como el motor del proyecto, ya que al utilizar esta iniciativa nuestro desarrollo se vuelve más sencillo y práctico. Es la mejor alternativa para unir todas las fuentes de información sin violentar su independencia de trabajo, a la vez que ayudan a construir un sistema común centralizado que sirva para todos los usuarios.

Es tanto el impulso ocasionado por OAI, que ahora se está utilizando el protocolo de recolección de metadatos en aplicaciones nunca pensadas, como por ejemplo el proyecto *Open Citation Project - Citebase* (Hitchcock et al, 2002)(<http://citebase.eprints.org/>), el cual es un proveedor de servicios que recolecta toda la información contenida

en varios proveedores de datos, cuando se realiza una búsqueda el sistema muestra los resultados ordenados por factor de impacto, adicionalmente permite navegar por los distintos artículos a través de sus referencias bibliográficas. Cuando muestra un registro (artículo) específico, nos brinda como complemento la siguiente información:

- ◆ Un gráfico que indica el nivel de citación y relevancia del artículo.
- ◆ Una lista de referencias de los documentos citados por el artículo.
- ◆ Una lista *Top* de los principales 5 documentos que citan este artículo, ordenados por factor de impacto.
- ◆ La lista de todos los artículos que citan este artículo y su ubicación actual.
- ◆ Una Lista *Top* de los 5 artículos más citados junto a este artículo (co-citados).
- ◆ La lista de todos los documentos citados junto a este artículo (co-citados).

Otro resultado destacado es el sistema *Paracite* (<http://paracite.eprints.org/>), el cual consiste en un servicio que, partiendo de una referencia bibliográfica, es capaz de buscar el documento en la red y si lo encuentra establecer los enlaces correspondientes.

En estos ejemplos anteriores, vemos como ayudan en la evolución de los sistemas de recuperación de información, ya que brindan nuevas herramientas para obtener resultados científicos y académicos más precisos y relevantes, a partir de una evaluación práctica de los metadatos; otra forma de utilización novedosa consiste en el intercambio de metadatos de documentos diferentes a los *e-prints*, como son el intercambio de información entre archivos históricos institucionales mediante el esquema de metadatos EAD, y la posibilidad de interoperar sistemas de almacenamiento de objetos de

aprendizaje, que manejan los modelos SCORM y LOM.

El uso del protocolo de recolección de metadatos, enmarcado en la iniciativa de archivos abiertos se está expandiendo cada día más, a tal punto que Google en su portal académico (<http://scholar.google.com/>) está indexando todos los proveedores de datos que trabajen esta iniciativa. Sin embargo, en Colombia no se ha visto el interés de nuestras bibliotecas por utilizar esta iniciativa, si revisamos los listados de participantes en el mundo (<http://archives.eprints.org/index.php?action=browse>), podemos observar que apenas tenemos registrados tres repositorios, de los cuales solo uno ha surgido como iniciativa de una bi-

blioteca, no hemos visto la necesidad de estar presentes en la escena mundial, por creer erróneamente en la complejidad que involucra la implantación de esta iniciativa. Sin embargo, gracias a la evolución de las tecnologías de la información y la comunicación, se nos da la oportunidad perfecta para que empecemos a trabajar en equipo, construyendo repositorios y bibliotecas digitales interoperables, intercambiando experiencias y desarrollos tecnológicos que nos posicionen como uno de los países líderes en manejo de la información digital en Latinoamérica, de lo contrario estaríamos condenados a permanecer atrasados, y continuar desapareciendo a ritmo acelerado de la escena académica y científica internacional.

## BIBLIOGRAFÍA

- Alamos Laboratory. El Laboratorio Nacional de Los Álamos principal centro de investigación de los Estados Unidos. <http://www.lanl.gov/>
- Ariel, H. «Creative Commons y la profundidad del copyright». *Revista Enredando* 353- 105º de la 4ª versión. [http://www.sindominio.net/afe/dos\\_copyleft/cc.pdf](http://www.sindominio.net/afe/dos_copyleft/cc.pdf).
- Arranz, D. *Apuntes de XML*. <http://programacion.com/tutorial/apuntesxml/>
- ARXIV. Archivo y servicio de impresos electrónicos en los campos de la Física, Matemáticas y Computación. <http://arxiv.org>
- Barrueco, JM. y I. Subirats. «Open archives initiative. Protocol for metadata harvesting (OAI-PMH): descripción, funciones y aplicaciones de un protocolo». *El profesional de la información* 12. 2. (2003): 99-106.
- Bergman, MK. «The deep Web: surfacing hidden value». *Journal of Electronic Publishing* 17. Agosto 2001. <http://www.press.umich.edu/jep/07-01/bergman.html>
- Berners-Lee, T. «Uniform Resource Identifiers (URI): Generic Syntax (1998)». <http://www.ietf.org/rfc/rfc2396.txt>
- Cole, T W. «Using OAI: innovations in the sharing of information». *Library Hi Tech* 21. 2. (2003): 115-117. <http://www.emeraldinsight.com/0737-8831.htm>
- Cole, T.; W. Mischo y T. Habing. «Introduction to the open archives initiative Protocol for metadata harvesting, introduction tutorial given at the ACM/IEEE Joint». Conference on Digital Libraries, Houston, TX, (2003). <http://dli.grainger.uiuc.edu/publications/TWcole/>
- Ginsparg, P.; R. Luce y H. Van De Sompel. *First meeting of the Open Archives initiative*. Octubre 29<sup>th</sup>, 1999. <http://www.openarchives.org/news/ups1-press.htm>.
- Gómez, ND. y OM. Arias. El cambio de paradigma en la comunicación científica. Información, cultura y sociedad, (2002). <http://eprints.rclis.org/archive/00003649/>
- GOOGLE. INC. Search report 2002 Year-End Google Zeitgeist. (<http://www.google.com/press/zeitgeist2002.html>)
- Hitchcock, S., *et al.* «Open Citation Linking: The Way Forward». *D-Lib Magazine*. 8. 10. (2002): <http://www.dlib.org/dlib/october02/hitchcock/10hitchcock.html>.
- Hunter, P. y M. Guy. «Metadata for Harvesting: The Open Archives Initiative, and how to find things on the Web». *The Electronic Library* 22. 2. (2004): 168-174. <http://www.emeraldinsight.com/0264-0473.htm>
- Lagoze, C., *et al.* The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0, Open Archives Initiative, June 2002. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- - -. «Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting: Specification and XML Schema for the OAI Identifier Format» <http://www.openarchives.org/OAI/2.0/guidelines-oai-identifier.htm>
- Lawrence, S. «Online or Invisible?». *Nature* 411. 6837. (2001): 521. <http://www.neci.nec.com/~lawrence/papers/online-nature01/>

- Lawrence, S. «Free online availability substantially increases a paper's impact». *Nature Web-debates* 31. Mayo 2001 <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- Lynch, CA. «Metadata harvesting and the Open Archives Initiative». *ARL Bimonthly Report*. 217. (2001):1-9. <http://www.arl.org/newsltr/217/mhp.html>
- Martínez Méndez, FJ. Blog Recuperación de Información en la Web. <http://irsWeb.blogspot.com/2005/02/indizacin-por-semantica-latente-en.html>
- Melero, R., et al. «Plataforma Digital de revistas científicas electrónicas españolas: Relación con el movimiento Open Access». VI Coloquio Internacional de Geocrítica - El impacto social y espacial de las nuevas tecnologías de la información y la comunicación; Barcelona Mayo 2004. <http://www.ub.es/geocrit/sn/sn-170-74.htm>
- Müller, U. «Tutorial: OAI and OAI-PMH for Beginners, An introduction to the Open Archives Initiative and the Protocol for Metadata Harvesting» *2nd OAForum workshop*. Lisbon. 5th-7th December, 2002 [http://www.oaforum.org/otherfiles/lisb\\_tutorial.ppt](http://www.oaforum.org/otherfiles/lisb_tutorial.ppt)
- Piñeiro, I. y E. Marmonti. «Metadatos una introducción». II jornadas sobre la Biblioteca Digital Universitaria <http://www.econ.uba.ar/www/servicios/Biblioteca/ponencias/metadatos%20marmonti%202.ppt>
- SEDIC. Grupo de Trabajo sobre normalización para la recuperación de información en Internet, Portal NormaWeb. <http://www.z3950.info>
- Suleman, H., et al. «Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 1: Mission and Progress, and Part 2: Services and Research». *D-Lib Magazine* 7. 9 (2001): <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html>
- Triska, R. y L. Cafe. «Open Archives: the Brazilian Digital Library subproject». *Ci. Inf.* 30. 3. (2001) : 92-96 <http://www.scielo.br/pdf/ci/v30n3/7291.pdf>
- Van De Sompel, H. y C. Lagoze. «The Santa Fé Convention of the Open Archives Initiative». *D-Lib Magazine* 6. 2. (2000): <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>
- - ; T., Krichel, M. L. Nelson, et al.» The UPS Prototype: An Experimental End-User Service across E-Print Archives». *D-Lib Magazine* 6. 2. (2000): <http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html>
- - -. Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting: Specification for an OAI Static Repository and an OAI Static Repository Gateway, Protocol Version 2.0 of 2002-06-14: <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>.
- WU MING 1, *Il Mucchio selvaggio*, 526. 25 de Marzo de 2003: [http://www.wumingfoundation.com/italiano/outtakes/copyleft\\_explicado.html](http://www.wumingfoundation.com/italiano/outtakes/copyleft_explicado.html)

